Proposed 2020 Census Data Products for Detailed Racial and Ethnic Groups

Nicholas Jones
Director of Race Ethnicity Research & Outreach
Population Division

National Advisory Committee on Racial, Ethnic, and Other Populations – Fall 2021 Virtual Meeting November 4-5, 2021



Background

- Census Bureau has a long history of tabulating detailed race and ethnicity group data and protecting information provided by respondents
- Over time, more and more granular census data have been published
- Advances in data science, more powerful computers, and externally accessible 'big data' has increased the risk of identifying individuals from published statistics
- To mitigate, the Census Bureau is implementing new disclosure avoidance methods for the 2020 Census
- Differentially private Disclosure Avoidance Systems (DAS) are being developed to protect 2020 Census data products for detailed race and ethnicity groups, tribes, and their sex and age characteristics
- Census Bureau's goal is to meet data user needs while implementing new disclosure avoidance methods for 2020 Census data products





Proposed 2020 Census Detailed DHC

- Population counts for detailed race, ethnicity, and tribal groups
- Basic demographic and housing characteristics for sex and age, household and family type, and tenure

- Stakeholder feedback, combined with review of table access rates, and subject matter expert knowledge informed development of 2020 Census data products
- 2020 Census data products include legally mandated statistics that will be supported by formal privacy systems





Timeline for Developing Detailed-DHC

Spring 2020

- DEMO team introduced to Group II Differential Privacy and Tumult Labs researchers
- Developed analytical tools for evaluating privacy and accuracy balance for producing detailed race ethnicity statistics

Summer – Fall 2020

• DEMO researchers evaluated 2010 Census data and how best to achieve balance of privacy and accuracy

Fall 2020

- DEMO presented proposal and recommendation to Census DSEP for total population and sex by age statistics
- DSEP provided initial approval for DEMO parameters and epsilon levels

Winter - Summer 2021

 DEMO put Detailed DHC work on hold to review 2020 Census data files for producing apportionment and redistricting data products





DEMO SafeTEx Analyses with Tumult Labs

Tumult Labs' SafeTEx tool allows Census Bureau researchers to assess interaction and balance of privacy and accuracy for producing detailed race, ethnicity, sex, and age data for the 2020 Census

Parameters for privacy and accuracy levels can be set to produce statistics on thousands of population groups

SafeTEx answers the question –

Given specific parameter settings, what data could we publish within an acceptable level of accuracy?



2020 Census Race and Ethnicity Code List

Appendix F.

Hispanic Origin and Race Code List

WHITE	1000-1999
White White (Checkbox) White	1000-1009 1000 1001
EUROPEAN (EXCEPT SPANISH)	1010-1749
Albanian	1010-1014
Albanian	1010
Alsatian	1015-1019
Alsatian	1015
Andorran	1020-1024
Andorran	1020
Armenian Armenian	1025-1029 1025
Austrian	1030-1034
Austrian	1030
Tyrolean	1031
Azerbaijani	1035-1039
Azerbaijani	1035
Basque	1040-1044
Basque	1040
French Basque	1041





Detailed Race and Ethnicity Groups

Hispanic or Latino
Asian
Black or African American
Native Hawaiian or Other Pacific Islander
White (Middle Eastern or North African)
White (European, Other)

Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY21-035.

* Sources: 2010 Census SF-2 data, 2016 ACS 5-year data

Argentinean
Bolivian
Chilean
Chilean
Colombian
Costa Rican
Cuban
Ecuadorian
Gustemalan
Honduran
Mexican
Nicarsguan
Panamanian
Paraguayan
Perevian
Puerto Rican
Spaniard
Uruguayan
Venezuelan

Bhutance
Burmere
Cambodian
Chinese, except Taiwar
Filipino
Hmong
Indonesian
Ivo Jiman
Japanere
Korean
Laotian
Malajvian
Malajvian
Malajvian
Malajvian
Malajvian
Nepolere
Ohinawan
Pakistani
Singaporean
Singaporean
Taiwanere
Thai

iligerian
kwandan
enegalese
ierra Leonean
omali
outh African
t Lucia Islander
t Maarten Islander
udanese
wasiland
anstanian
ogo
riniiddaln/Tobagonian
urks and Caicos Islander
Igandan
ilis Virgin Islander
incent-Grendine Islande
/act Indian
ambian
imbabwean

British Virgin Islander Burkinabe

ominica Islande

Sambian.

irenadian

esotho

Madagascan

Guadeloupe Islander

Carolinian
Chuukeze
Fijian
Gusmanian or Chamorro
I-Kiribati
Kozraean
Marisan Islander
Marshallese
Native Hawsiian
Mi-Yanustu
Palusun
Papus New Guinean
Pohnpeian
Suipaneze
Samoan
Solomon Islander
Tokelauan

Assyrian
Bahraini
Berber
Chaldean
Egyptian
Emirati
Iranian
Iranja
Israeli
Jordanian
Kurdish
Kuwaiti
Lebanese
Libyan
Moroccan
Omani
Palestinian
Gatari
Saudi

Liechtensteine Lithuanian Luxembourge Maltese Moldovan Monegasque Montenegrin New Zealander Norwegian Pennsulvania Germai Polish Portuguese Roma Romanian Russian Scandinavia Scots-Irish Scottish Serbian Slovak Swedish Tatar Turkish Ukrainian

Albanian Andorran

Armenian

Azerbaijan Basque

Belgian Belgian Bosnian and Herzegovinian

British Bulgarian

Canadian Carpatho Rusyn

Croatian Cypriot Czech Czechoslovakian Danish Dutch

English Estonian Faroe Islander Finnish Finno Ugrian

French Canadiar Frisian

Georgian

German Greek Greenlande

Lapp

Latvian

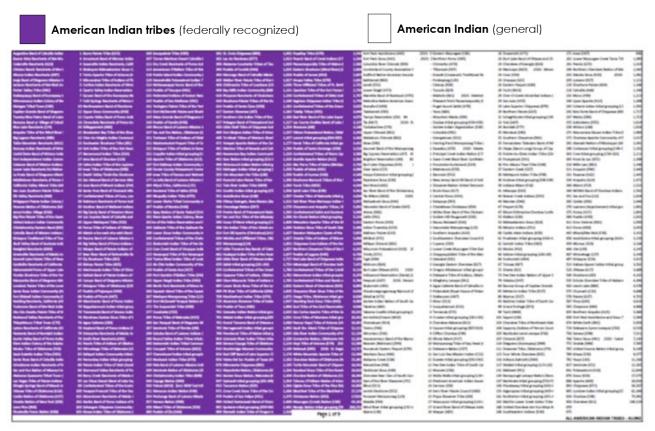
Shape your future START HERE >

2020CENSUS.GOV





Detailed American Indian Tribes and Alaska Native Village Groups



^{*} Source: 2010 Census CPH-T-6 data



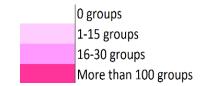
Shape your future START HERE >

Alaska Native (general)



Number of Detailed Groups UNDER Established Threshold at Various Levels

Threshold	White (n=95)	Black (n=63)	Asio (n=2		NHPI (n=20)		A (n=	Hispanic (n=20)	
Illieshold									
	Total	Total	Alone	AOIC	Alone	AOIC	Alone	AOIC	Total
100	0	0	2	1	2	1	490	453	0
200	0	1	2	2	6	2	613	573	0
300	1	4	2	2	8	3	690	652	0
400	3	5	2	2	8	4	753	710	0
500	3	5	2	2	10	6	793	754	0
600	6	6	2	2	11	7	833	786	0
700	6	10	2	2	11	7	867	817	0
800	6	13	2	2	11	7	894	836	0
900	6	13	2	2	11	7	910	860	0
1,000	6	14	2	2	12	9	930	880	0
1,500	10	20	2	2	13	11	981	945	0
2,000	11	22	2	2	13	11	1,017	990	0
2,500	12	23	2	2	13	12	1,053	1,018	0
3,000	13	26	3	2	14	12	1,080	1,049	0
3,500	13	26	4	2	15	12	1,099	1,064	0
4,000	13	27	4	2	15	12	1,112	1,080	0
4,500	13	30	4	2	15	13	1,120	1,104	0
5,000	15	30	4	2	16	13	1,133	1,113	0



* Sources: 2010 Census Asian Brief, 2010 Census Pacific Islander Brief, 2010 Census SF-2 data, 2010 Census CPH-T-6, 2016 ACS 5-year data

Shape your future START HERE >



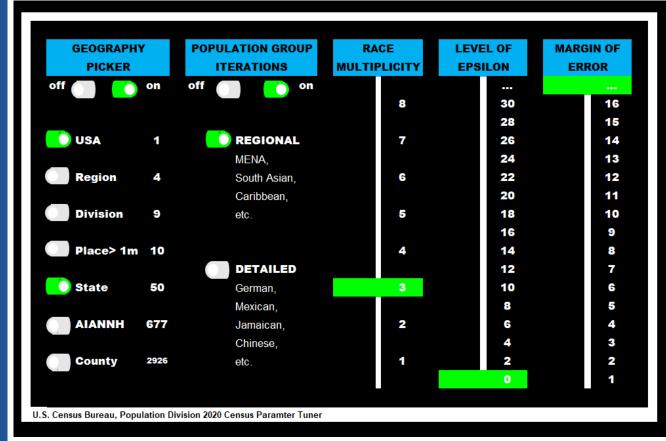
DEMO Research Goals

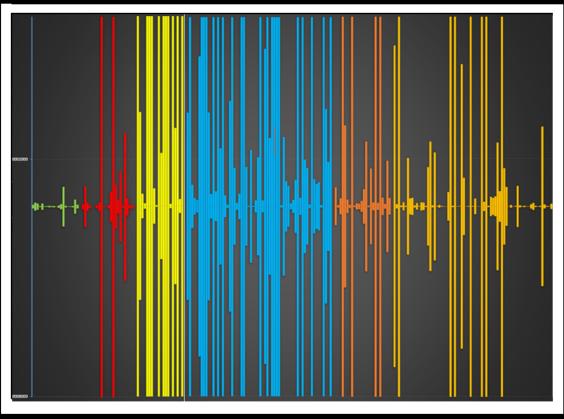
Develop parameter settings for algorithm to produce differentially private 2020 Census data for detailed race and ethnicity groups, and their sex and age characteristics

- Explore how to establish privacy and accuracy levels to produce statistics for our nation's myriad detailed racial and ethnic population groups
- Allow for equity across all major race and ethnicity categories
- Provide same amount, or more data, than provided from 2010 Census



Parameter Tuner and Noise Visualization





Hispanic or Latino
Asian
Black or African American
Native Hawaiian or Other Pacific Islander
White (Middle Eastern or North African)
White (European, Other)

Shape your future START HERE >



2020 Census Detailed DHC Table 1: Population Counts

Proposed 2	Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells								
2020	Cell Count	Indent	Table Shell						
Table #	Cell Count	indent	Table Stell						
T1.		0	TOTAL POPULATION [1]						
T1.		0	Universe: Total population						



2020 Census Detailed DHC Table 2: Sex by Age

Proposed	Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells														
2020 Table #	Cell Count	Indent	Table Shell	ı											
T2.		C	SEX BY AGE	E FOR SELEC	TED AGE C	ATEGORIES [49]								
T2.		C	Universe:	Total popu	ation										
T2.	1	C	Total:												
T2.	2	1	L	Male:											
T2.	3	2	2		Under 5 y	ears									
T2.	4	2	2		5 to 9 year	rs									
T2.	5	2	2		10 to 14 y	ears									
T2.	6	2	2		15 to 17 y	ears									
T2.	7	2	2		18 and 19	years									
T2.	8	2	2		20 years										
T2.	9	2	2		21 years										
T2.	10	2	<u>)</u>		22 to 24 y	ears									
T2.	11	2	<u>)</u>		25 to 29 y	ears									
T2.	12	2)		30 to 34 y	ears									
T2.	13	2)		35 to 39 y	ears									
T2.	14	2)		40 to 44 y	ears									
T2.	15	2)		45 to 49 y	ears									
T2.	16	2	2		50 to 54 y	ears									
T2.	17	2)		55 to 59 y	ears									



2020 Census Detailed DHC Table 3: Household Type

Proposed	roposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells								
2020 Table #	Cell Count	Indent	Table Shell						
T3.		0	HOUSEHOLD TYPE [9]						
T3.		0	Universe: Household	S					
T3.	1	0	Total:						
T3.	2	1	Family ho	useholds:					
T3.	3	2		Married co	uple family				
T3.	4	2		Other famil	y:				
T3.	5	3			Male house	eholder, no	spouse present		
T3.	6	3			Female hou	useholder, r	no spouse present		
T3.	7	1	Nonfamily	households:	:				
T3.	8	2		Householde	er living alo	ne			
T3.	9	2		Householde	er not living	g alone			
			NOTE: When a catego	ory other than	n Total Popi	ulation is se	lected, the data in this table refer to the race, Hispanic or Latino origin, or tribe of the householder.		



2020 Census Detailed DHC Table 4: Tenure

Proposed 2	Proposed 2020 Census Detailed Demographic and Housing Characteristics File (Detailed DHC) Table Shells																			
2020 Table #	Cell Count	Indent	Table Shell	Shell																
T4.		0	TENURE [4]																	
T4.		0	Universe: Occup	Universe: Occupied housing units																
T4.	1	0	Total:																	
T4.	2	1	Owi	ned with	a mortgag	e or a loan														
T4.	3	1	Owi	ned free a	and clear															
T4.	4	1	Ren	ter occup	oied															
			NOTE: When a	category	other thar	n Total Pop	ulation is se	elected	l, the da	ta in thi	is table i	refer to	the race	, Hispani	or Latin	o origin,	or tribe	e of the h	ouseholde	er.



Data Product Comparisons: 2010 vs 2020

Topics	2010	2020
Number of Population Tables	57	4
Number of Geographies	104	4
Threshold for National Sex by Age Characteristics	100	500
Threshold for National Total Count forAIAN Tribes and VillagesDetailed Race and Ethnicity Groups	0 100	10 10
Number of Race and Ethnicity Groups (not inclusive of AIAN tribes and villages)	67	334
Disclosure Avoidance Technique	Swapping	Differential Privacy

Note: Estimates based on 2010 Census SF-2 data, 2016 ACS 5-Year data, and 2020 Census proposal for Group II data products.



Next Steps for Developing Detailed DHC

Where Are We Now?

Fall 2021...

- DEMO researchers re-engaging with R&M and Tumult after completion of 2010 Census PL release
- Engaging with data users and public on Detailed DHC proposal and use cases
- Evaluate privacy/accuracy tradeoff for the Detailed DHC data

Before "final" parameter setting and epsilon decisions are made...

- Discuss proposed 2020 Census Group II data product plans with tribal leaders in formal AIAN Tribal Consultations
- Address feedback and recommendations from Census Bureau Advisory Committees
- Assess implications that post-processing will have on production of detailed race and ethnicity data products





Questions for NAC

What questions or concerns do you have about any potential changes to the DHC and/or the Detailed DHC?

Are there any critical decennial census statistics that are not currently proposed in the 2020 Data Products crosswalk and/or geographic detail that you require?

What levels of geography (e.g., places, school districts) are needed for detailed race ethnicity population groups, and for what purpose, policy, program, etc.?



